

WIE VIELE PARONYMPAARE GIBT ES EIGENTLICH?

DAS ZUSAMMENSPIEL AUS KORPUSLINGUISTISCHEN UND REDAKTIONELLEN VERFAHREN ZUR ERMITTLUNG EINER PARONYMSTICHWORTLISTE

Der Autor ist wissenschaftlicher Mitarbeiter der Abteilung Lexik am Institut für Deutsche Sprache in Mannheim.

Ausgangslage – Voraussetzungen

Die Stichwortliste bildet eine tragende Säule von gedruckten als auch von Online-Wörterbüchern. Daher ist die Konzeption und Erarbeitung einer Stichwortliste auch eine zentrale Aufgabe in der Anfangsphase des am IDS in Angriff genommenen Paronymwörterbuchs (vgl. Storzjohann 2014). Welche Überlegungen dabei angestellt, welche Ideen entwickelt und welche Pläne bzw. Entscheidungen schließlich umgesetzt wurden, soll im Folgenden nachvollzogen werden.

Der Wortbestand einer Sprache geht rein quantitativ prinzipiell gegen unendlich. Die Erstellung einer Liste einzelner (Stich-)Wörter, die man in ein Wörterbuch aufnehmen sollte, kommt insofern stets einer mehr oder minder großen Auswahl gleich. Diese Auswahl erfolgt unter bestimmten methodischen Vorgaben. Deren Abstimmung auf die Zielvorgaben, die reflektierte,

stringente Anwendung und sorgfältige Dokumentation wird spätestens dann einsichtig, wenn man mit der Frage konfrontiert wird: Warum taucht dieses eigentlich doch völlig überflüssige Wort in der Stichwortliste auf, während jenes doch viel wichtigere fehlt? Eine Antwort kann nur im Hinweis auf die angewandten methodischen Grundsätze bestehen: Da eine Stichwortliste eine Auswahl ist, eine Auswahl jedoch häufig subjektiv gefärbt erscheint und willkürlich anmutet, gilt es, diese Auswahl auf wissenschaftlich-konzeptionellem Wege zu objektivieren, transparent zu gestalten und sie dadurch zu legitimieren. Für die Erarbeitung einer Stichwortliste bedeutet das, dass der Weg (die Methodik und deren Nachvollziehbarkeit) mindestens ebenso wichtig ist wie das Ziel (die eigentliche Stichwortliste).

BEI DER ERSTELLUNG EINER STICHWORTLISTE GILT ES, UNTER BESTIMMTEN METHODISCHEN VORGABEN EINE AUSWAHL AUS DEM WORTBESTAND EINER SPRACHE ZU TREFFEN

Ebenso trivial wie notwendig ist beispielsweise die Beantwortung folgender Fragen:

- Welcher Wortschatz soll ausgewählt werden, also welche Quellen bzw. Korpora können als Auswahlbasis dienen? Dies hängt maßgeblich vom Wörterbuchtyp ab, also davon, ob man ein fachsprachliches, historisches oder allgemeinsprachliches Wörterbuch erarbeiten will.
- Was soll rein formal gelistet werden? Das allgemeinsprachliche Wörterbuch beispielsweise behandelt überwiegend einzelne Wörter, man kann aber auch Wortgruppen, z. B. Phraseologismen, zum lexikografischen Gegenstand machen.
- In welcher Form soll das Stichwort angegeben werden? Als sog. Nennform fungiert in der Regel die Grundform, also beispielsweise der Infinitiv bei Verben, der Nominativ Singular bei Substantiven, der Positiv bei Adjektiven.

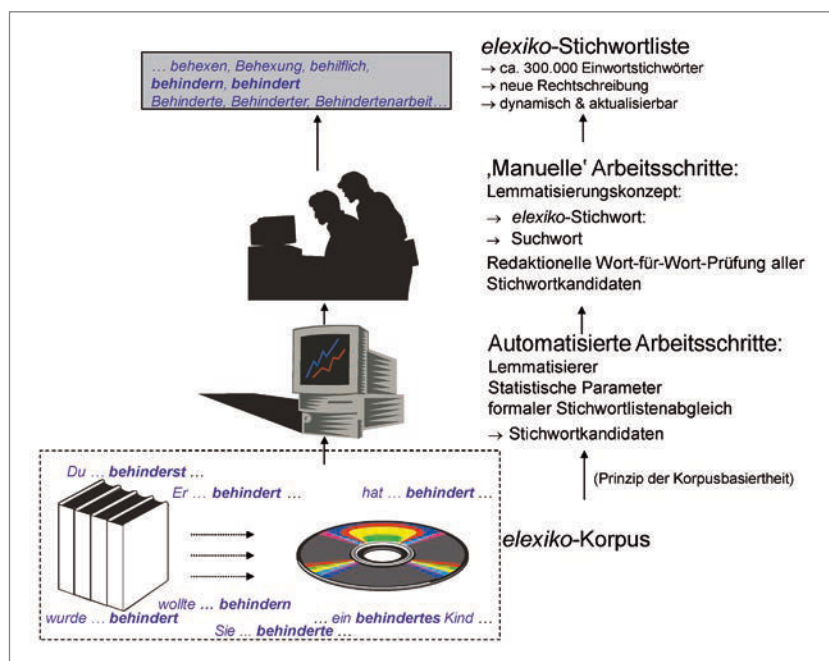


Abb. 1: Die Erarbeitung der *elexiko*-Stichwortliste. Geringfügig abgeänderte und erweiterte grafische Darstellung aus Schnörch (2005, S. 88)

Welche Überlegungen anzustellen sind, lässt sich auch am Beispiel der Erstellung einer korpusbasierten Stichwortliste für das allgemeinsprachliche elektronische Wörterbuch *lexiko* nachvollziehen (vgl. Schnörch 2005) (siehe Abb. 1).

Die etwas nostalgisch anmutende Grafik (s. Abb. 1) visualisiert die zentralen Arbeitsschritte, die sich im Falle der *lexiko*-Stichwortliste auch in der Phase der Wörterbucherstellung bewährt haben. Gleichzeitig wird damit in gewisser Weise auch die Traditionslinie angedeutet, in der das Paronymwörterbuch zu verorten ist. Vorteilhaft ist es dabei, auf die grundsätzlichen Erfahrungen bei der kompletten Neubearbeitung einer Stichwortliste auf Korpusbasis zurückgreifen zu können: Für die Neuerstellung einer Stichwortliste auf Korpusbasis ist methodisch die Verzahnung redaktioneller und automatisierter Arbeitsschritte unabdingbar. Es gilt, die Erfordernisse des spezifischen Wörterbuchtyps bzw. Wörterbuchs und die der Online-Publikation zu berücksichtigen.

DIE ANGESTREBTE DYNAMISCHE STICHWORTLISTE IST DAS RESULTAT EINER METHODISCHEN VERKNÜPFUNG VON REDAKTIONELLEN UND AUTOMATISCHEN ARBEITSPROZESSEN

Den Rahmen gibt die Funktion des Stichworts vor, z. B. als Überschrift eines Wortartikels, Suchwort, Struktur-/Sortierelement u. Ä., das Ziel ist das Konzept einer dynamischen Stichwortliste, das die Streichung und Neuaufnahme von Stichwörtern im laufenden Arbeitsprozess in begründeten Fällen unterstützt. Aufgrund der skizzierten Vorgeschichte ist es nicht nötig,

hier solche allgemeinen Fragestellungen nochmals zu erörtern. Stattdessen soll im Folgenden die Aufmerksamkeit auf einige Aspekte gelenkt werden, die sich ganz speziell mit der Erstellung einer Stichwortliste für ein Paronymwörterbuch ergeben. Dies gestaltet sich an einigen Punkten komplexer, ganz konkrete Unterschiede sind:

Bei *lexiko* entschied allein die Korpusfrequenz der lemmatisierten und redaktionell geprüften Wörter über die Aufnahme in die Stichwortliste, Auswahlbeschränkungen inhaltlicher o. ä. Art gibt es in einem allgemeinsprachlichen Wörterbuch grundsätzlich nicht. Demgegenüber suchen Nachschlagende in einem Paronymwörterbuch ausschließlich Paronyme bzw. Paronympaare. Anders als bei *lexiko* sollte daher aus methodischer Sicht ein weiterer Filter, nämlich eine definitorische Eingrenzung der Korpusrecherche vorausgeschaltet werden, um von vornherein möglichst viele, bestenfalls nur Paronyme (genauer: Paronympaare

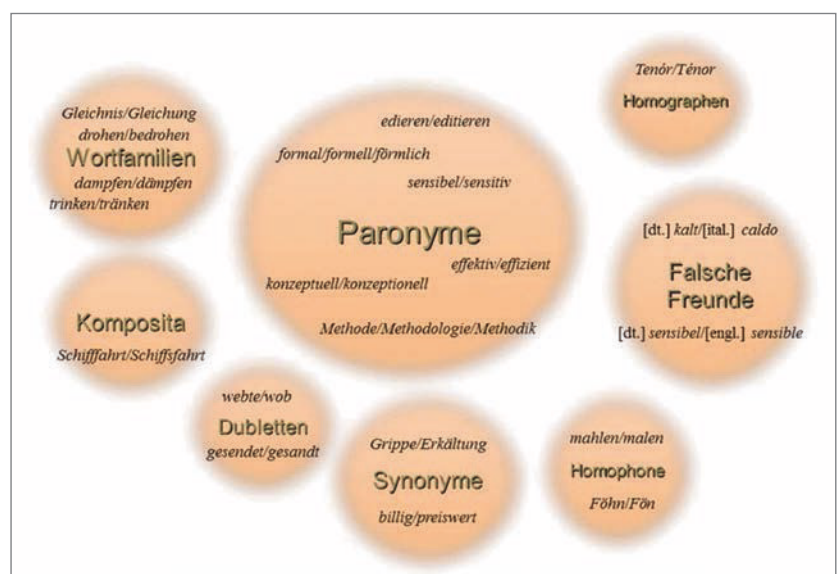


Abb. 2: Paronymie und andere Relationen mit Verwechslungspotenzial (häufig aufgrund von formaler Ähnlichkeit)

bzw. -gruppen!) korpusbasiert und automatisch zu ermitteln. Hieraus ergibt sich jedoch ein Dilemma: Von begrifflicher Seite ist die Paronymie definitorisch bislang nicht eindeutig festgelegt. Eine Ursache dafür ist, dass sich ganz unterschiedliche Forschungsrichtungen diesem Phänomen mit teilweise voneinander abweichenden Definitionsansätzen nähern (vgl. detaillierter dazu Lăzărescu 1999, bes. Kap. 2).

DER PARONYMIEBEGRIFF IST DEFINITORISCH BISLANG NICHT EINDEUTIG FESTGELEGT, WODURCH SICH SCHWIERIGKEITEN BEI DER ERMITTLUNG VON PARONYMPAAREN BZW. -GRUPPEN ERGEBEN

Daher scheint es zum gegenwärtigen Zeitpunkt ratsam, mit einer möglichst unvoreingenommenen, eher unscharfen Vorstellung an die Aufgabe der Erstellung der Stichwortlisten heranzutreten und zunächst von lediglich zwei zentralen Definitionskriterien auszugehen: Verwechslungspotenzial in Verbindung mit formaler Ähnlichkeit. Diese Komponenten spielen jedoch nicht nur bei der Paronymie eine Rolle, wie Abb. 2 (siehe S. 17) verdeutlicht.

Obschon Kriterien wie die formale Ähnlichkeit und das Verwechslungspotenzial zentrale Bausteine für die Definition von Paronymie darstellen, lassen sich aus dieser Zusammenstellung auch weitere ableiten: beispielsweise der Bezug auf EINE Sprache (im Gegensatz zu den sog. falschen Freunden), die Bedeutungsähnlichkeit UND Bedeutungsverschiedenheit (im Gegensatz zur Synonymie), der Bezug auf EINE Wortart (im Gegensatz zur Wortfamilie) u. a. Fasst man dies zusammen, so lassen sich die beiden zentralen Kriterien präzisieren bzw. zusammen mit weiteren für eine vorläufige Arbeitsdefinition bündeln (vgl. zur weiteren Annäherung an das Phänomen der Paronymie u. a. Lăzărescu 1999, Storjohann 2014, Storjohann/Schnörch i. Vorb.):

Paronyme

- kommen innerhalb einer Sprache vor,
- gehören meist einer Wortart an,
- sind sich in Schreibung und Aussprache sehr ähnlich,
- bestehen aus einer gemeinsamen (morphologischen) Invariablen (oft die etymologische Wurzel) und jeweils unterschiedlichen (morphologischen) Variablen (oft verschiedene Suffixe),
- besitzen einen semantischen Überschneidungsreich (können beispielsweise teilsynonym sein).

Aufgrund dieser Ausgangslage wird die erste Phase bei der Erarbeitung einer Stichwortliste für das Paronymwörterbuch von zwei Komponenten geprägt, nämlich a) dem Kriterienbündel einer vorläufigen Arbeitsdefinition sowie b) einer Sammlung bereits analytisch und lexikografisch erprobter, so gesehen muster-gültiger Beispiele wie *effizient/effektiv/effektivvoll*, *sensibel/sensitiv*, *formell/formal* usw. Diese beiden Komponenten dienen im größeren Zusammenhang der Theoriebildung und der empirischen Stützung und stehen in einem engen Wechselspiel miteinander: Der Definitionsansatz liefert über die bloße Intuition hinausgehende Vorgaben für die weitere Beispielauswahl. Jedes weitere Beispiel (und dessen lexikografische Analyse) hilft bei der Präzisierung der Definition, was den Horizont auf weitere Beispiele erweitert usw.

WAS SIND PARONYME? IM FOKUS SOLL DAS VERWECHSLUNGSPOTENZIAL IN VERBINDUNG MIT FORMALER ÄHNLICHKEIT STEHEN

Das Fundament

Die eben skizzierte Annäherung an den Paronymbegriff von zwei Seiten legt für die Auswahl der Stichwortkandidaten des geplanten Paronymwörterbuchs in der Anlaufphase konsequenterweise auch ein zweigleisiges Vorgehen nahe: Zunächst wurden Paronymkandidaten im wahrsten Sinne des Wortes gesammelt. Einzelne Projektmitarbeiterinnen und -mitarbeiter trugen von der oben vorgestellten Arbeitsdefinition geleitet entsprechendes Material zusammen, das ihnen bei

Paronym A	Paronym B	Paronym C	Paronym D	Basis	VariantenteilA	VariantenteilB	VariantenteilC
abstrus	absurd			abs	trus	urd	
Adaption	Adaptation			Adapt	ion	ation	
Alternation	Alteration			Alter	nation	ation	
assoziativ	assoziiert			assozi	ativ	iert	
authentisieren	authentifizieren			authenti	sieren	fizieren	
autoritär	autoritativ			autorit	är	ativ	
begnadigt	begnadet			begnad	igt	et	
bildhaft	bildlich			bild	haft	lich	
desparat	disparat			sparat	de	di	
Diskurs	Diskussion			Disku	rs	ssion	
Duo	Duett			Du	o	ett	
effektiv	effizient	effektiv		eff	ektiv	izient	ektiv
Effizienz	Effektivität			Eff	izienz	ektivität	
Egozentrik	Egoismus			Ego	zentrik	ismus	
egozentrisch	egoistisch			ego	zentrisch	istisch	
Elektrik	Elektronik			Elektr	ik	onik	
elektrisch	elektronisch			elektr	isch	onisch	
emotional	emotionell			emotion	al	ell	
empfindsam	empfindlich			empfind	sam	lich	
erstmalig	erstmalig			erstmal	s	ig	
ethisch	ethnisch			eth	isch	nisch	
experimental	experimentell			experiment	al	ell	
faschistisch	faschistoid			faschist	isch	oid	
feudal	feudalistisch			feudal	Ø	istisch	
fiktiv	fiktional			fikti	v	onal	
föderal	föderalistisch			föderal	Ø	istisch	
Fossil	Fossilie			Fossil	Ø	ie	
fremdsprachig	fremdsprachlich			fremdsprach	ig	lich	
fundamental	fundamentalistisch			fundamental	Ø	istisch	
funktional	funktionell			funktion	al	ell	
geistig	geistlich			geist	ig	lich	
generational	generationell			generation	al	ell	
Genese	Genesis			Genes	e	is	
genial	genialisch	genialistisch		genial	Ø	isch	istisch
Geriat	Gerontologe			Ger	iat	ontologe	
geriatrisch	gerontologisch			ger	iatrisch	ontologisch	
Gigantismus	Gigantomanie			Gigant	ismus	omanie	

Tab. 1: Ausschnitt aus der Datenbank mit den Stichwortkandidaten (Materialsammlung)

der Lektüre auffiel, ihnen von Dritten mitgeteilt wurde usw. Diese eher assoziativ, jedenfalls wenig systematisch ermittelte Sammlung wurde anschließend durch Material ergänzt, das aus Wörterbüchern stammt, die Stichwörter unter dem Aspekt der Verwechselbarkeit behandeln. Die herangezogenen Wörterbücher waren Pollmann/Wolk (2010), Duden 9 und Müller (1973). Eine Subgruppe dieser Gesamtmenge sind mutmaßlich wiederum Paronyme (vgl. dazu nochmals Abb. 2); diese wurden extrahiert. Auf diese Weise entstand eine Datenbank mit insgesamt 277 Datensätzen (vgl. zu Illustrationszwecken Tab. 1). Das entspricht einer ebensolchen Anzahl potenzieller Paronympaare bzw. -gruppen, denn neben 258 Zweierpaaren (z. B. *elektrisch/elektronisch*, *legal/legitim*, *parteiisch/parteilich*) finden sich darunter auch 15 Dreiergruppen (z. B. *original/originell/originär*, *mehrfach/mehrmals/mehrmalig*, *Methode/Methodik/Methodologie*) und drei Vierergruppen (z. B. *anomal/anormal/abnormal/unnormal*).

DIE AUSWERTUNG DER STICHWORTKANDIDATEN (MATERIALSAMMLUNG) ERGIBT, DASS DIESE IN BEZUG AUF FORMALE ÄHNLICHKEITEN UND UNTERSCHIEDE SYSTEMATISIERT WERDEN KÖNNEN

Ein Blick in diese Materialsammlung mit 277 potenziellen Stichwortpaaren bzw. -gruppen für ein Paronymwörterbuch offenbart eine auffällige Eigenschaft: Die Kandidaten bilden kein Sammelsurium willkürlicher Paarungen. Segmentiert man die Zeichenketten, so erkennt man, dass die Paarungsbeziehungen in formaler (morphologischer) Hinsicht auf muster- und regelhaften Zusammenhängen basieren. Die Stichwortkandidaten sind demnach in Bezug auf formale Ähnlichkeiten (invariable Zeichenkette, z. B. Stamm) und Unterschiede (variable Zeichenketten, z. B. Suffixe) systematisierbar. Die Gruppierbarkeit nach Mustern eröffnet gute Möglichkeiten für den korpusgestützten und systematischen Ausbau der Kandidatenliste von Paronymen.

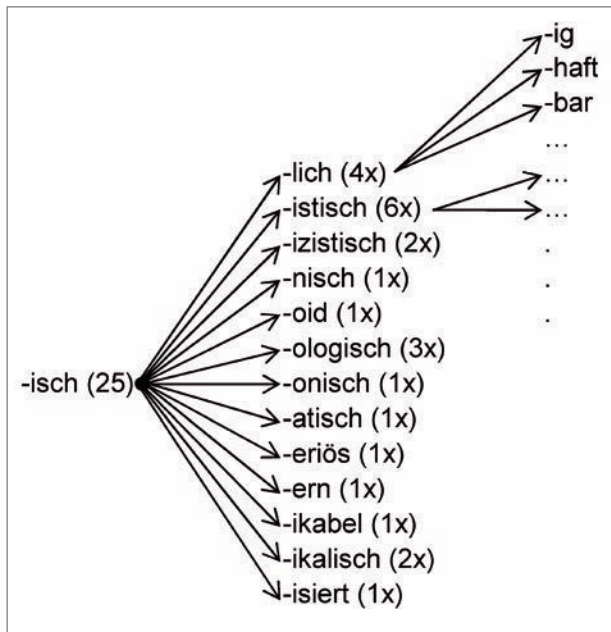


Abb. 3: Variable *-isch* und ihre Alternativformen

Ausbau der Stichwortliste auf Korpusbasis: Die systematische Materialerhebung

Für die Analyse und Gruppierung müssen die Paronympaarungen der Materialsammlung zunächst in identische (invariable) und sich unterscheidende (variable) Bestandteile segmentiert werden. Diese Zerlegung der Zeichenketten kann, muss sich aber nicht zwangsläufig mit den Ergebnissen einer morphologischen Analyse decken: Das Ziel der Paronymsegmentierung ist es auch nicht, beispielsweise Erkenntnisse in Bezug auf die zugrunde liegenden Wortbildungsstrukturen zu gewinnen. Im Hinblick auf die geplante weitere computergestützte Recherche in Korpora ist es hier eine notwendige Voraussetzung, die Zeichenketten so zu zerlegen, dass das Material Muster und Regelmäßigkeiten zu erkennen gibt, auf deren Basis Prozeduren computergestützter Korpusrecherche- und -analyse umgesetzt werden können.

DIE SEGMENTIERUNG DER PARONYMPAARE IN ZEICHENKETTEN ERFOLGT NICHT, UM WORTBILDUNGSSTRUKTUREN ERSCHLIESSEN ZU KÖNNEN, SONDERN UM MUSTER UND REGELHAFTIGKEITEN ZU ERKENNEN

Die Vielfalt und Vielzahl der ermittelten Muster überrascht allenfalls auf den ersten Blick. In einem Zwischenschritt wurden 176 Paronympaare /-gruppen stichprobenartig in Bezug auf die Musterbildung und -verteilung getestet. Die häufigste Form stellt mit 147

Fällen die Variation am Wortende dar (z. B. *kompl -ex/-iziert, prakti -sch/-kabel, Appli -kation/-zierung*). Daneben finden sich 15 Fälle von Variation am Wortanfang (z. B. *un-/de- motiviert, un-/a- sozial*), sieben Fälle mit einer Variation in der Wortmitte (z. B. *unbestreitbar/unbestritten*) sowie einige Kombinationen dieser drei Haupttypen (z. B. *scheinbar/anscheinend, hölzern/holzig, männlich/mannhaft*). Im Falle der Endvariation bewegt sich deren Frequenz in der Materialsammlung zwischen sehr häufig (z. B. 22x für *-al/-ell*), häufig (11x *-ig/-lich*) oder selten (z. B. 1x für *-al/-iert*).

Die Muster bei der Variation sind freilich nicht immer als simple, d. h. systemhaft gleich bleibende 1:1-Korrespondenzen realisiert, nach dem Motto: Unterschiedliche invariable Bestandteile in Kombination mit den immer gleich bleibenden beiden Variablen ergeben eine Reihe von Paronympaarungen; die Variablen bilden ihrerseits vielmehr ein stark verästeltes System alternierender bzw. konkurrierender Formen, und damit Paarungen bzw. Paarungsmöglichkeiten, was die Systematisierung erschwert. Am eindrucksvollsten ist dieses Phänomen vielleicht am Beispiel von *-isch* am Wortende zu illustrieren (vgl. Abb. 3). Die 25 auf *-isch* endenden Paronympartner besitzen 13 Entsprechungen, mit denen ihrerseits wiederum mehrere konkurrierende Variablen korrespondieren können. Alternierende Variablen am Wortbeginn sind beispielsweise *un-*, *de-*, *di-*, *a-*.

Insgesamt ließen sich schließlich in der vollständigen Materialsammlung mit 277 Paronympaarungen aufgrund der vorgenommenen Segmentierung und Gruppierung 179 unterschiedliche Muster, also Kombinationen eines variablen mit einem invariablen Bestandteil, isolieren. Die Daten dieser Muster wurden an Rainer Perkuhn vom Programmbereich Korpuslinguistik des IDS weitergeleitet. Er überprüfte, welche und wie viele Realisierungen diese Muster in einer auf DEREKO basierenden Grundformenliste mit 326.946 Einträgen (DEREWO) erfahren.

DER MUSTERABGLEICH ERGIBT 9.246 KANDIDATEN FÜR PARONYMPAARUNGEN

Jeder der so ermittelten Paronymkandidaten wurde mit einer relationalen Häufigkeitsangabe versehen; diese und andere bereits eingefügten statistischen Angaben können wichtige Parameter für die spätere Binnenstrukturierung der Stichwortliste (z. B. in Häufigkeitsklassen) darstellen.¹ Der Musterabgleich ergab in Summe 9.246 konkrete paarweise Realisierungen aller potenziellen Paronymmuster, d. h. 9.246 Kandidaten für Paronympaarungen.

Beim Blick in die entsprechende Liste springt natürlich eine ganze Reihe von sehr produktiven morphologischen Mustern ins Auge. Neben dieser formalen Eigenheit reicht die Bandbreite aller Paarungen, wenn man (rein synchron) auch die inhaltliche Seite berücksichtigt, von mutmaßlich echten Paronymkandidaten über diverse Zweifelsfälle bis hin zu Analogbildungen z. B. aus grammatischen Paradigmen und kuriosen Zufälligkeiten ohne jegliches Verwechslungspotenzial, z. B.

- regelhaft: Infinitive/Partizipien:
laden/geladen, mahlen/gemahlen (neben *brauchen/gebrauchen*)
- zufällige Analogien:
Danke/Gedanke, Flügel/Geflügel, Stirn/Gestirn vs. *Rede/Gerede*;

- (kuriose) Zufälligkeiten:
[-abe/-ebung] *Schwabe/Schwebung* (neben *Eingabe/Eingebung*)
[-iv/-orisch] *kurativ/kuratorisch, definitiv/definitorisch* (neben *provokativ/provokatorisch*);
[-ieren/-alisieren] *signieren/signalisieren* (neben *minimieren/minimalisieren*);
[-heit/-zahl] *Torheit/Torzahl* (neben *Minderheit/Minderzahl*);
[-al/-iert] *normal/normiert, rational/rationiert* (neben *proportional/proportioniert*);
[-al/-isch] *dialektal/dialektisch* (neben *patriarchal/patriarchisch*);
[-sam/-fältig] *einsam/einfältig* (neben *sorgsam/sorgfältig*).

Vor allem dieser Umstand macht eine redaktionelle Durchsicht der automatisch generierten Daten erforderlich, bei der besonders inhaltliche Gesichtspunkte zu berücksichtigen sind, v. a. das Verwechslungspotenzial.

Redaktionelle Bearbeitung und Strukturierung der Stichwortkandidatenliste: Materialorganisation und -auswahl

Das Ziel der redaktionellen Bearbeitung ist es, möglichst viele, in erster Linie inhaltlich nicht als Paronymkandidaten einzustufende Paarungen aus der computergenerierten Stichwortkandidatenliste zu entfernen. Um dieses Ziel zu erreichen, fiel die Entscheidung zugunsten einer sehr einfachen Methode: Drei Mitglieder des Projektteams kennzeichneten unabhängig voneinander zunächst alle 9.246 Kandidatenpaare in einer Datenbankspalte entweder mit ‚ja‘ (= kann meiner Einschätzung nach zu Verwechslungen bzw. Verwendungsunsicherheiten im Sinne von Paronymie kommen) oder ‚nein‘ (= kann meiner Einschätzung nach nicht zu Verwechslungen bzw. Verwendungsunsicherheiten im Sinne von Paronymie kommen). Auf dem Wege des Mehrheitsvotums konnten dann alle Paronymkandidaten der computergenerierten Liste einer von drei Kategorien zugeordnet werden:

3x ja/2x ja+1x nein = potenzielles Paronympaar fürs Paronymwörterbuch,
 1-0x ja bzw. 2-3x nein = kein Paronympaar, für weitere Analysen (aktuell) uninteressant.

Nach diesem Indizierungs- und Kategorisierungsschritt ergibt die quantitative Auswertung eine Summe von 1.936 Paronympaaren, die mehrheitlich (= 3x bzw. 2x) ein ‚ja‘ erhielten. Diese Mehrheitsfälle bieten sich für eine Bearbeitung im Paronymwörterbuch an.

**DIE COMPUTERGENERIERTEN KANDIDATEN
 WERDEN AUF IHRE INHALTLICHE
 BEZIEHUNG HIN ÜBERPRÜFT UND GGF.
 AUS DER LISTE ENTFERNT: 1.936
 PARONYMPAARE FÜR DIE BEARBEITUNG
 VERBLEIBEN**

Mit einer solch hohen Zahl war zu Beginn der Arbeiten wohl kaum zu rechnen. Umso notwendiger und vorteilhafter ist es, diese Menge nochmals untergliedern zu können. Im Zuge der mustergeleiteten, automatischen Extraktion von Paronympaaren (s. o.) wurde zu jedem ermittelten Partnerwort dessen relative Häufigkeitsklasse verzeichnet, wobei der kleinste Wert die größte (relative!) Frequenz markiert. Auf dieser Basis wurden vier Kategorien abgeleitet, sog. ‚Prioritätsklassen‘. Darunter werden Teilmengen verstanden, die die Paronymstichwortliste unter dem Gesichtspunkt der Frequenz für die lexikografische Arbeit strukturieren. Für die Klassifikation wurde folgende Verfahrensweise umgesetzt und es wurden dabei zwei Grundsatzfälle unterschieden:

- Gehören beide Paronympartner einer Häufigkeitsklasse an, die eine Prioritätsklasse bilden, so werden sie eben dieser Prioritätsklasse zugeordnet und entsprechend mit I, II, III oder IV markiert.
- Differieren die jeweils ermittelten Häufigkeitsklassen zweier Paronyme so stark, dass diese unterschiedlichen Prioritätsklassen zugeordnet werden, so entscheidet jener Partner mit der höheren Häu-

figkeitsklasse darüber, in welche Prioritätsklasse das Paronympaar eingeordnet wird; dies wird dann mit den Zwischenstufen I,5, II,5, III,5, IV,5 und V,5 markiert. Wenn beispielsweise Paronym A die Häufigkeitsklasse 8 besitzt und folglich der Prioritätsklasse I zugeordnet wird, Paronym B hingegen 16 und III, dann wird das Paar mit I,5 markiert.

Häufigkeitsklasse	Prioritätsklasse
8-11	I/I,5
12-15	II/II,5
16-19	III/III,5
20-Rest	IV/IV,5

Tab. 2: Zuordnung von Häufigkeitsklassen und Prioritätsklassen der Stichwortparonympaare

Auf die Prioritätsklasse I (= sehr häufig) fallen 109 Paronymkandidaten, auf II 553, auf III 799, auf IV 460. Im Zuge der Hierarchisierung der Paronymkandidaten nach Prioritätsklassen wurden diese auch in Bezug auf die Wortart markiert.

Neben den Häufigkeitsklassen kann bei den Paronymkandidaten noch deren reihenhafte Konstruktionsweise (Invariable bzw. Stamm + verschiedene Variablen/Affixe) zur Subkategorisierung herangezogen werden. Alphabetisch sortiert lassen sich in der Liste der Stichwortkandidaten zahlreiche Paare zu größeren Gruppen zusammenfassen und hierarchisieren: Paare mit Negation können unnegierten Paaren zur Seite gestellt werden, und Zusammensetzungen können als Nebengruppen einer Hauptgruppe mit Simplicia zugeordnet werden. Eine weitere Reihe von Paaren lässt sich zu Mustern zusammenfassen, wobei dann das häufigste als Haupteintrag fungiert. Unter den Simplicia *Technik/Technologie* (vgl. Tab. 3) sowie unter dem frequentesten Paar *damals/damalg* (vgl. Tab. 4) werden beispielsweise folgende Nebenpaare subsummiert:

Stamm	Paronym A	ParonymB	Muster	möglicher Haupteintrag	Priorität auf Basis von Häufigkeitsklassen
Agrartechn	Agrartechnik	Agrartechnologie	ik_ologie	Technik-Technologie	3,5
Antriebstechn	Antriebstechnik	Antriebstechnologie	ik_ologie	Technik-Technologie	3,5
Atomtechn	Atomtechnik	Atomtechnologie	ik_ologie	Technik-Technologie	3
Basistechn	Basistechnik	Basistechnologie	ik_ologie	Technik-Technologie	4
Batterietechn	Batterietechnik	Batterietechnologie	ik_ologie	Technik-Technologie	4
Bautechn	Bautechnik	Bautechnologie	ik_ologie	Technik-Technologie	3,5
Biotechn	Biotechnik	Biotechnologie	ik_ologie	Technik-Technologie	2,5
Breitbandtechn	Breitbandtechnik	Breitbandtechnologie	ik_ologie	Technik-Technologie	4
Brennstoffzellen-techn	Brennstoffzellen-technik	Brennstoffzellen-technologie	ik_ologie	Technik-Technologie	4
Chiptechn	Chiptechnik	Chiptechnologie	ik_ologie	Technik-Technologie	4
Computertechn	Computertechnik	Computertechnologie	ik_ologie	Technik-Technologie	3
Dieseltechn	Dieseltechnik	Dieseltechnologie	ik_ologie	Technik-Technologie	4
Digitaltechn	Digitaltechnik	Digitaltechnologie	ik_ologie	Technik-Technologie	3,5
Energiespartechn	Energiespar-technik	Energiespartechnologie	ik_ologie	Technik-Technologie	4
Energietechn	Energietechnik	Energietechnologie	ik_ologie	Technik-Technologie	3
Spitzentechn	Spitzentechnik	Spitzentechnologie	ik_ologie	Technik-Technologie	3,5
Techn	Technik	Technologie	ik_ologie	Technik-Technologie	1,5

Tab 3: Haupt- und Nebenpaarungen am Beispiel des Hauptpaares *Technik/Technologie* (Ausschnitt)

Stamm	Paronym A	ParonymB	Muster	Endergebnis	Selektion	möglicher Haupteintrag	Priorität auf Basis von Häufigkeitsklassen
abermal	abermals	abermalig	-s -ig	ja	NP	damals-damalg	2,5
damal	damals	damalg	-s -ig	ja	HP	damals-damalg	1
ehemal	ehemals	ehemalg	-s -ig	ja	NP	damals-damalg	1,5
einstmal	einstmals	einstmalig	-s -ig	ja	NP	damals-damalg	2,5
erstmal	erstmals	erstmalig	-s -ig	ja	NP	damals-damalg	1
letztmal	letztmals	letztmalig	-s -ig	ja	NP	damals-damalg	2,5
mehrmal	mehrmals	mehrmalg	-s -ig	ja	NP	damals-damalg	2
nachmal	nachmals	nachmalig	-s -ig	ja	NP	damals-damalg	3,5
nochmal	nochmals	nochmalig	-s -ig	ja	NP	damals-damalg	1,5
oftmal	oftmals	oftmalig	-s -ig	ja	NP	damals-damalg	2,5
vormal	vormals	vormalig	-s -ig	ja	NP	damals-damalg	2

Tab 4: Haupt- und Nebenpaarungen am Beispiel des Hauptpaares *damals/damalg*

259 Hauptpaarungen (vergleichbar dem Typ *Technik/Techologie, damals/damalig*) sowie 537 Nebenpaarungen (vergleichbar dem Typ *Antriebstechnik/Antriebstechnologie, Atomtechnik/Atomtechnologie, ehemals/ehemalig, oftmals/oftmalig* usw.) ergeben insgesamt 796 Paronympaare, zusätzlich nach Häufigkeit gegliedert.

HÄUFIGKEITSKLASSE UND KONSTRUKTIONSWEISE (INVARIABLE BZW. STAMM UND VERSCHIEDENE VARIABLEN / AFFIXE) BILDEN SUBKATEGORIEN

Am Ende der skizzierten Bearbeitungsschritte lässt sich nun aus der angelegten Gesamtdatenbank eine Auswahl von korpusbasiert ermittelten, frequentativ geschichteten und nach Musterhaftigkeit (i. w. S. morphologischen Gesichtspunkten) subkategorisierten Stichwortkandidaten für das Paronymwörterbuch extrahieren: 259 Hauptgruppen sowie 537 Nebengruppen, die diesen zugeordnet werden können.

MATERIAL- BZW. KORPUSGRUNDLAGE UND STICHWORTANSATZ MÜSSEN GEKLÄRT UND ANSCHLIESSEND AUF DEN SPEZIFISCHEN WÖRTERBUCHTYP ABGESTIMMT WERDEN

Dass für den einen oder die andere bei dieser Auswahl immer auch ein Paronympaar fehlen wird, ist erwartbar, warum es fehlt, sollte jedoch nachvollziehbar gemacht worden sein – ergänzbar (beispielsweise aus einer anderen Teilmenge der Stichwortkandidatenliste überführbar) sollte ein Paronympaar aufgrund des dynamischen Stichwortkonzepts ebenfalls sein.

Zusammenfassung und Fazit

Für die Erarbeitung einer Stichwortliste egal welchen Wörterbuchtyps müssen zunächst einige grundsätzliche Fragen, beispielsweise die Material- bzw. Korpusgrundlage und der Stichwortansatz geklärt werden.

Bei allen Gemeinsamkeiten gilt es jedoch, ab einem gewissen Stadium das Vorgehen auf den Wörterbuchtyp abzustimmen, so auch im Falle des Paronymwörterbuchs (im Gegensatz etwa zum allgemeinsprachlich ausgerichteten *lexiko*-Wörterbuch). Auf einer abstrakteren Ebene sind die Besonderheiten, die das methodische Vorgehen beim Erstellen der Stichwortliste des Paronymwörterbuchs maßgeblich beeinflussen, die folgenden:

- Die Orientierung an der Definition des Untersuchungsgegenstands:
Ein Paronymwörterbuch sollte selbstredend ausschließlich Paronyme enthalten. Es ist ein Problem, aber in diesem Fall auch eine Chance, dass bisherige Versuche, Paronyme zu definieren, meist stark strukturalistisch-kategoriell geprägt sind und das Phänomen nicht korpusgestützt bzw. empirisch untersucht wurde. Begriffsbestimmungen werden darüber hinaus aus ganz unterschiedlichen Perspektiven vorgenommen, wobei der Fokus auf eine Einzelsprache gerichtet oder die kontrastive Ausrichtung (Mutter- vs. Fremdsprache) betont wird. Das führt zwangsläufig zu definitorischen Differenzen. Für die Stichwortliste des Paronymwörterbuchs besteht das Dilemma darin, etwas auswählen zu müssen (Paronympaare), von dem man eine unscharfe Vorstellung hat, da ganz unterschiedliche begriffliche Umschreibungen existieren. Methodisch gesehen ist insofern der Weg (die Erarbeitung eines Paronymwörterbuchs) auch das Ziel (Schärfung der Paronymdefinition mit Fertigstellung des Wörterbuchs).
- Methodenabstimmung und Auswahl der Stichwortpaare:
Bei der Stichwortliste für das Paronymwörterbuch handelt es sich um eine Auswahl (ebenso stellt auch eine noch so große Stichwortliste beispielsweise all-

gemeinsprachiger Wörterbücher eine Auswahl dar). Daher ist es nötig und sinnvoll, die Methoden, die zu dieser Auswahl führen, auf das Ziel abzustimmen, sie reflektiert und stringent anzuwenden sowie nachvollziehbar zu protokollieren. Um diese Auswahl im konkreten Fall durchzuführen, wurden in der skizzierten Weise eine Reihe automatischer, korpusbasierter Verfahren mit redaktionellen Auswertungsschritten verzahnt.

– Binnenstrukturierung:

Da Wörterbücher online nicht von A bis Z publiziert werden, macht es wenig Sinn, die Auswahlmenge an Stichwörtern alphabetisch zu sortieren. Eine binnenstrukturierte Stichwortliste ist in der Praxis jedoch sehr von Vorteil, z. B. um Teilmengen für die Bearbeitung zu ermitteln. Im Falle von korpusbasiert ermittelten Stichwortkandidatenlisten können dazu Kriterien wie Frequenz und Streuung herangezogen werden. Bei den Paronymen scheint es darüber hinaus sinnvoll, auch formstrukturelle, i. w. S. morphologische Zusammengehörigkeiten zu berücksichtigen (vgl. Haupt- vs. Nebengruppen). Ob später noch weitere Parameter hilfreich sind, muss ggf. noch geprüft werden; einzubeziehen wäre z. B. die Angabe der Ähnlichkeit von Paronymen in Bezug auf den Überschneidungsgrad ihrer Kollokationspartner.

– Dynamisches Stichwortlistenkonzept:

Die computergestützt ermittelte Liste von Stichwortkandidaten für das Paronymwörterbuch ist von drei ProjektmitarbeiterInnen unabhängig voneinander redaktionell geprüft worden. Sie ist insofern deutlich objektiver als die von *lexiko*, deren Stichwortkandidaten von lediglich einer Person dahingehend überprüft wurde, ob sie dem Lemmatisierungskonzept entsprechend den Status eines *lexiko*-Stichwortes haben oder nicht. Dadurch ist es gelungen, das Defizit zu kompensieren, das der Paronymbegriff als Auswahlhilfe aufgrund seiner

definitorischen Unschärfe bzw. Mehrdeutigkeit besitzt. Die Stichwortliste fungiert auch nicht nur als Pool für die lexikografische Abarbeitung. Sie ergibt darüber hinaus eine Menge von Wortpaaren/-gruppen für die empirische Untersuchung des Phänomens Paronymie (mittels der semantisch-analytischen Untersuchung). Die Stichwortliste bildet so gesehen sowohl eine tragende Säule des Paronymwörterbuchs als auch das Gerüst zur Erforschung der Paronymie. Um dieser Doppelfunktion gerecht werden zu können, ist das Konzept einer dynamischen, d. h. für Veränderungen offenen, Stichwortliste Voraussetzung.

– Stichwortpräsentation:

Im Paronymwörterbuch stehen (mindestens) zwei Stichwörter und deren Vergleich im Fokus. Dieser kontrastive Ansatz stellt weniger für die Präsentation der Stichwörter als für die der Wörterbuchinformationen eine Herausforderung dar. Diese anzugehen heißt vor allem, im Rahmen der elektronischen, korpusbasierten Lexikografie sich einer weiteren zentralen Aufgabe zu stellen.

DIE KONTRASTIVE PRÄSENTATION DER WÖRTERBUCHINFORMATIONEN IST HERAUSFORDERUNG UND ZIEL DES PARONYMWÖRTERBUCHS

Die Auswahl und Strukturierung der Stichwortliste ist fürs Erste abgeschlossen und in den Grundzügen dokumentiert, mit anderen Worten: Die Stichwortliste, eine Säule des Paronymwörterbuchs, steht zur Verfügung.

Literatur:

- Belica, Cyril (2001ff.): Kookkurrenzdatenbank CCDB – V3.3. Eine korpuslinguistische Denk- und Experimentierplattform. Institut für Deutsche Sprache: Mannheim. <<http://corpora.ids-mannheim.de/ccdb/>> (Stand: 26.8.2015).
- CCDB: <<http://corpora.ids-mannheim.de/ccdb/>> (Stand: 26.8.2015), s. a. Belica 2001ff.
- DEReKo: Deutsches Referenzkorpus, <www1.ids-mannheim.de/kl/projekte/korpora/> (Stand: 26.08.2015).
- DEReWo: Grundformenliste, <www1.ids-mannheim.de/kl/projekte/methoden/derewo.html> (Stand: 26.08.2015).
- Duden 9. Richtiges und gutes Deutsch (2007): Wörterbuch der sprachlichen Zweifelsfälle. Hrsg. von der Dudenredaktion, bearb. v. Peter Eisenberg, unter Mitwirkung von Franziska Münzberg, Kathrin Kunkel-Razum. 6. Aufl. Mannheim u. a.
- Lăzărescu, Ioan (1999): Die Paronymie als lexikalisches Phänomen und die Paronomasie als Stilfigur im Deutschen. Bukarest: Editura Anima.
- Müller, Wolfgang (1973): Leicht verwechselbare Wörter. Duden-Taschenwörterbücher, Bd. 17. Mannheim: Bibliographisches Institut.
- Pollmann, Christoph / Wolk, Ulrike (2010): Wörterbuch der verwechselten Wörter. Stuttgart: Pons.
- Schnörch, Ulrich (2005): Die *ellexiko*-Stichwortliste. In: Haß, Ulrike (Hg.): Grundfragen der elektronischen Lexikographie. *ellexiko* – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter, S. 71-90.
- Storjohann, Petra (2014): Das künftige Projekt „*ellexiko*: Paronymwörterbuch“. In: SPRACHREPORT 1/2014, S. 22-29.
- Storjohann, Petra / Schnörch, Ulrich (i.Vorb.): Verwechslungsfehler oder Wandel? Sprachliche Veränderungen der Gegenwart und ihre lexikografische Dokumentation. In: Themenheft „Lexikografie“, Marie Vachková, Martin Šemeík, Věra Kloudová (Hg.): Germanistica Pragensia, AUC Philologica. Bd. XXIV.

Anmerkungen

- ¹ Dafür besten Dank an Rainer Perkuhn! ■